

## Chapter 8

### Contemporary Approaches to Modelling the Consumer

Debbie Isobel Keeling  
Professor of Marketing  
University of Sussex

#### 1. Introduction

Modelling plays an important role in consumer psychology and is used to understand consumers across diverse contexts. For example, modelling of consumer profiles, loyalty and switching behaviours can support marketing decisions around the launch of new brands. Consumer psychologists design campaigns to encourage healthy living routines based on modelling of consumer attitudes, peer influences and individual differences in personality and health status. Others use modelling to effectively segment their consumers into identifiable groups as a means of tailoring marketing communications, product offerings and recommendation systems. Further, modelling is fundamental to understanding consumer decision-making, such as understanding risk-taking behaviours amongst consumers in investments, health, travel and so forth. Whatever the application, **modelling** as a term in consumer psychology should be understood as a whole process, encompassing defined objectives, design, data collection and analysis, for the development and/or testing of theories relating to why consumers behave as they do. The result of this process is a '**model**' that can be expressed in multiple forms, including path models, equations, charts, matrices etc., which provides a holistic overview of the relationships between key influencers (variables) on consumer behaviour.

Why is modelling such an important process for understanding consumers? The generic value of modelling in consumer psychology is threefold. First, it provides vital insights into the complexity of the psychology of consumers' behaviours. There are multiple variables that influence consumers' behaviours, the role of modelling is to make sense of this complexity by identifying the most influential variables on behaviour and exploring how these variables work together to produce behavioural outcomes. Second, modelling offers as an outcome an integrative framework, namely a model, that both captures the relationships between variables and informs understanding of how to influence consumer behaviours. By identifying the most important variables and also the (combined) level and nature of their influence on behaviours, interventions can be designed that utilise these variables to optimum effect. Third, the resultant models offer insight into how to assess the success and/or impact of such interventions on consumers' behaviours. This is achieved both by identifying the key variables to measure and assess, and providing guidance on the expected level of impact of each variable on behaviour.

Whilst I have stressed the value of modelling in terms of understanding consumer behaviours, these behaviours can be diverse. For example, they can be direct observations of performance behaviours (e.g., sales, learning, fitness levels), use or consumption behaviours (e.g., purchase, repurchase, store visits, application of the product or experience) and choice behaviours (e.g., choosing between insurance policies, brand

preferences). Proxy measures in the form of self-reported behaviours or intentions are used when observations of actual behaviour are not readily available. Of course, consumer psychology models usually include some combination of emotional, cognitive and behavioural variables alongside social, demographic and geographic variables.

To demonstrate the three values of modelling, let us take a look at a classic conceptual model – the Theory of Planned Behaviour (TPB) devised by Ajzen (1991). Figure 1 shows the TPB in its simplest form. This is an example of how a model can be used as a visual representation of a theory, where the key influences on behaviour are indicated and the hypothesised relationships between those influencers represented. (*N.B.: see later for a note on conceptual versus statistical models.*) First, the TPB model identifies that four important variables (attitude towards the behaviour, subjective norm, perceived behavioural control, and intention) are influential on behaviour. Through ordering these variables, the model also indicates the relative relationships between these variables and with behaviour. Second, the model visually represents that three variables (attitude towards the behaviour, subjective norm and perceived behavioural control) shape intention, which subsequently shapes behaviour. Perceived behavioural control also has a direct impact on behaviour. From this model, we know that to influence behaviour we can narrow our interventions down to these four variables. For example, in campaigns to stop consumers smoking our interventions would not only require efforts to change consumer attitudes towards smoking, but also the social context and associated perceived norms, plus the individual's perceptions of their ability to give up smoking. Third, in tandem, we can also use these variables as a means of evaluating the success of our interventions and where they succeeded (or otherwise).

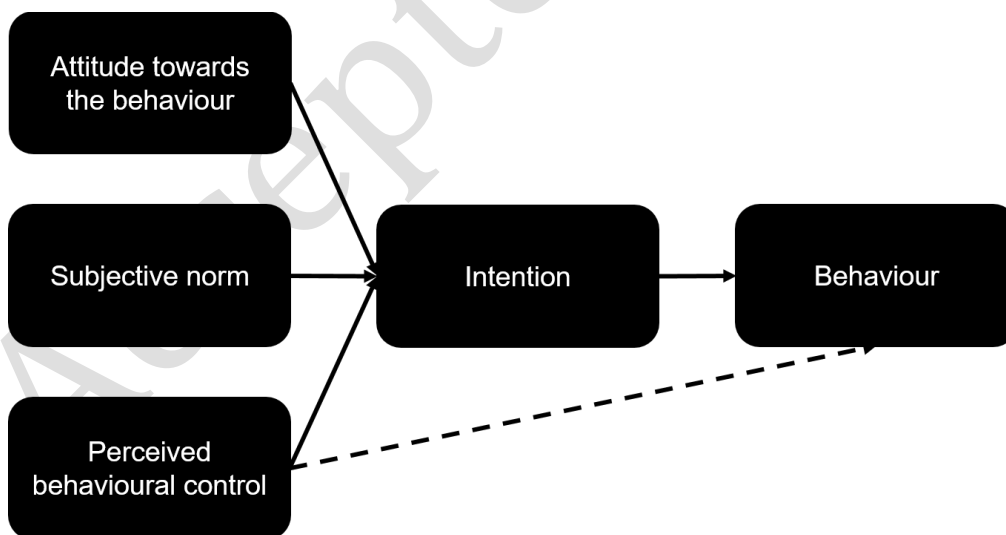


Figure 1: The (Simplified) Theory of Planned Behaviour (Ajzen 1991)

Contemporary developments in modelling have particularly focused on increasingly sophisticated methods of analysis and supporting software, including the increasing availability of specialist open access software. These developments in turn have

facilitated advancements in the specification of models aiding the description, prediction and explanation of consumer behaviour. In this chapter we will focus on the broad category of standard statistical modelling (mainly written from a frequentist perspective) popular within the field of consumer psychology. The aim of the chapter is to give you an overview of the underpinning principles and concepts that guide modelling. As such, the diverse goals and key building blocks of standard statistical modelling are explained. Offering two practical examples: technology adoption and the health consumer; segmentation and profiling in a B2B context, plus troubleshooting tips for using modelling in consumer psychology and identifying some of the statistical packages available to support you. Where relevant you are guided to more in-depth readings to explore specific analyses or issues. The chapter ends by reflecting on the future directions of modelling in consumer psychology and how these new directions will challenge and shape our thinking about the nature and application of modelling.

### 2. Learning Objectives

After reading this chapter, you should be able to:

- 1) Explain the standard statistical approach to modelling
- 2) Distinguish the diverse goals of standard statistical modelling
- 3) Understand the building blocks of this modelling approach
- 4) Address some of the pitfalls when using modelling in consumer psychology
- 5) Understand the scope of software on offer to help you with the modelling process
- 6) Identify the future directions of modelling in consumer psychology and how these are challenging our perspectives on modelling

### 3. Standard Statistical Modelling

Standard statistical modelling is one of the most widely applied families of modelling in consumer psychology and one in which there is constant innovation. The general aim of modelling from a statistical point of view is to develop an understanding of the relationships between important variables, and to represent these relationships within a defined framework (i.e., a model). This understanding is not limited to the bivariate relationships between 2 variables, but, most importantly, includes an exploration of the multivariate relationships between many variables. Multivariate relationships help us to understand how variables operate in combination with one another. Modelling aims to simplify this complexity through development of an organisational framework – the model. The resultant models can come in many forms, for example, graphs, equations, matrices, and figures. As such, this modelling process uses a diverse range of data inputs (e.g., transactional datasets, clickstream data, consumer surveys, laboratory or field experiments, and observations of behaviours) and analyses (e.g., regression, structural equations modelling, Bayesian approaches). Note that in this chapter we are using the term modelling in its widest sense to refer to a process of model development, and not in its narrow sense referring to specific types of analysis.

Typically, standard statistical modelling studies can be categorised as either exploratory

or confirmatory. In **exploratory** modelling we seek to identify and define possible relationships between the variables in which we utilise the data and the analytical method to help us to define the nature of those relationships. In **confirmatory** modelling we seek to test pre-specified relationships between our variables. As such, we must have a defined theory of how variables will work together and, therefore, which relationships will be significant. Exploratory and confirmatory modelling should be viewed as part of a model development continuum. During the modelling process, consumer psychologists often use multiple steps that move from an exploratory (development) basis in the early steps, to a confirmatory (testing) basis in later steps.

Thus the difference between exploratory and confirmatory analyses can be understood in terms of the role of modelling in relation to theory. A **theory** can be classically defined as: *“a set of interrelated constructs (concepts), definitions and propositions that present a systematic view of phenomena by specifying relationships among variables, with the purpose of explaining and predicting the phenomena”* (Kerlinger, 1986, p.9). In exploratory analyses the goal is to develop a theory, which we achieve through formulating and building a model. We may derive our model variables from one or a number of sources, for example, managerial intelligence and observations. Existing theory may play a role, but it is used in combination with other sources. Our model developed using exploratory analyses helps us to derive a number of propositions about our model variables and hence contribute to development of a theory. A **proposition** is a broad statement describing a relationship between two (or more) variables. In confirmatory analyses, the goal is to test our theory, achieved through testing our model. Hence theory underpins the specifically defined hypothesised and testable relationships between concepts. A **hypothesis**, takes propositions one step further, and formulates a more specific statement that is empirically testable. A hypothesis **operationalizes** the proposed relationship and puts it in an empirically testable form. In confirmatory analysis, by testing our model we are providing evidence for a proposed theory.

### ***3.1 Goals of Standard Statistical Modelling: Description, Prediction & Explanation***

Perhaps the most important goal of modelling is the ability to succinctly account for **relationship complexity**. That is, modelling helps us to both make sense of the multiple relationships that exist between variables and to then derive interventions to influence these variables. For example, healthcare professionals need a better understanding of what drives patient behaviour in order to work with patients on developing health promotion programmes. As such, we may want to understand how to encourage patients to avoid smoking or drinking too much alcohol. Of course, patient behaviours are influenced by more than one variable. As healthcare professionals we want to understand the most important influential variables and how they work together. For example, how do peers influence patients in social situations, such as, family parties or in restaurants? We could examine the bivariate relationships separately, but that would only provide a partial picture (often too simplistic) of the processes of consumer behaviour.

In examining relationship complexity, we can distinguish three specific goals of modelling: description; prediction; explanation.

#### ***The Goal of Description***

Modelling may be used as a means of summarizing or organising the key variables into a simplified structure, enabling us to describe the relationships between variables and consumer behaviour. This is most often used in the first steps of developing theory, that is, within exploratory modelling. The idea is not to suggest or test causality, but to provide a way of organising the potential relationships between variables and behaviour into an understandable pattern. That is, the goal is to capture the associations between the variables in a more 'usable' format. Regression techniques, often referred to as exploratory analytical techniques, are a common technique applied in the context of descriptive modelling.

*For example: As a first stage in developing a health promotion programme, we might first observe a set of consumers in different social settings focusing on their interaction with peers. From these observations we might derive a set of influential variables that impact on smoking and drinking behaviours and set these out in an exploratory model.*

### **The Goal of Prediction**

Modelling may be used to predict consumer behaviours given the input values of the independent variables. There are multiple analytical approaches to predictive modelling (e.g., PLS-SEM), all of which focus on predicting future behaviours. Predictive modelling sits at the intersection of exploratory and confirmatory modelling as it can contribute to both. Whilst it has been neglected as a technique to aid theory development, recent developments, especially in terms of analysis, have made this type of modelling both more accessible and more relevant to contemporary consumer psychology problems. For example, from an exploratory perspective, with the increasing availability of larger and richer datasets, predictive modelling techniques offer a way to develop theory by exploring the complex relationships and patterns that can be hard to initially observe and theorise through more traditional methods. This can be an excellent way of identifying and prioritising the best predictors of behaviour from a very large possible set. From a confirmatory perspective, predictive modelling offers an important test of developed models through their application to practice. By explicitly testing the 'confirmed' patterns between variables and behaviour in practice, predictive modelling can provide confidence in a model, allow comparisons between competing models and, in the case where models do not effectively predict behaviours, suggest improvements to existing explanatory models. That is, predictive modelling can be an important tool to establish the ecological validity of the model and its ongoing development.

*For example: As a second stage in developing a health promotion programme, we might seek a more parsimonious model in terms of the number of important variables. Using predictive modelling, from the set of variables identified in stage one, we can identify the most influential variables on our outcome behaviours and derive a simpler model.*

### **The Goal of Explanation**

Arguably the largest application of modelling within consumer psychology is the explanation of consumer behaviour. As such, a large body of consumer psychology

models fit within the realm of confirmatory modelling. That is, when the goal of modelling is explanation, we aim to test our hypotheses regarding how a variable or group of variables will shape behaviour. The justification for the relationships comes from the underpinning theory, which is clearly set out in a model. Through subsequent statistical techniques (such as SEM, referred to as confirmatory techniques), we seek to confirm (within statistical boundaries) whether variables ( $X_1$ ,  $X_2$  etc.) shape behaviour ( $Y$ ) in the way that was hypothesised. Advances in statistical analysis have led to the development of increasingly sophisticated techniques that allow us to more easily analyse complex relationship patterns, through simultaneously estimating several interrelated dependence relationships. It is not that these techniques did not exist before, rather that they have become more accessible and manageable for non-statisticians. Note that we should be very careful within consumer psychology to use the term 'causes', as we cannot often justify this (depending on the methodology used and context of the modelling).

*For example: As a third stage in developing a health promotion programme, we might seek to confirm our refined model by putting this to the test in a lab/field study. At this stage we have a defined set of variables that we can now measure and seek to confirm their influence on the outcome variables (e.g., smoking or drinking behaviours). We may test this through presenting stimuli in an experimental situation, or use a combination of questionnaire and observation data.*

#### 4. The Building Blocks of Models

Within modelling there is some specific terminology that you need to be aware of, referred to here as the key building blocks of models.

##### 4.1 Model Formats

Models come in various formats; the two most dominant formats in consumer psychology being path models and equations.

**Path Models:** Figure 2 is an example of a path model. A good **path model** should summarise the key variables and their hypothesised relationships with each other and with the dependent variable(s). Whilst figure 1 may be more properly classed as a conceptual model, in that it offers proposed relationships, figure 2 is more specific in that it indicates the relative value (e.g., nature, direction) of those relationships that we are able to statistically derive. Hayes (2013) makes an important distinction between the conceptual model (e.g., figure 1) and path models (e.g., figure 2), which represents the hypothesised relationships, and the **statistical model**, which represents the actual relationships being tested in an analysis (*see Hayes, 2013, for a detailed discussion*). You will also come across the terms **inner model** or **structural model** to refer specifically to the depiction of the key variables and their hypothesised relationships. The terms **outer model** or **measurement model** refers to relationships between the variables and indicator items (*that is, the items used to measure each variable – see 4.3 on measurement*).

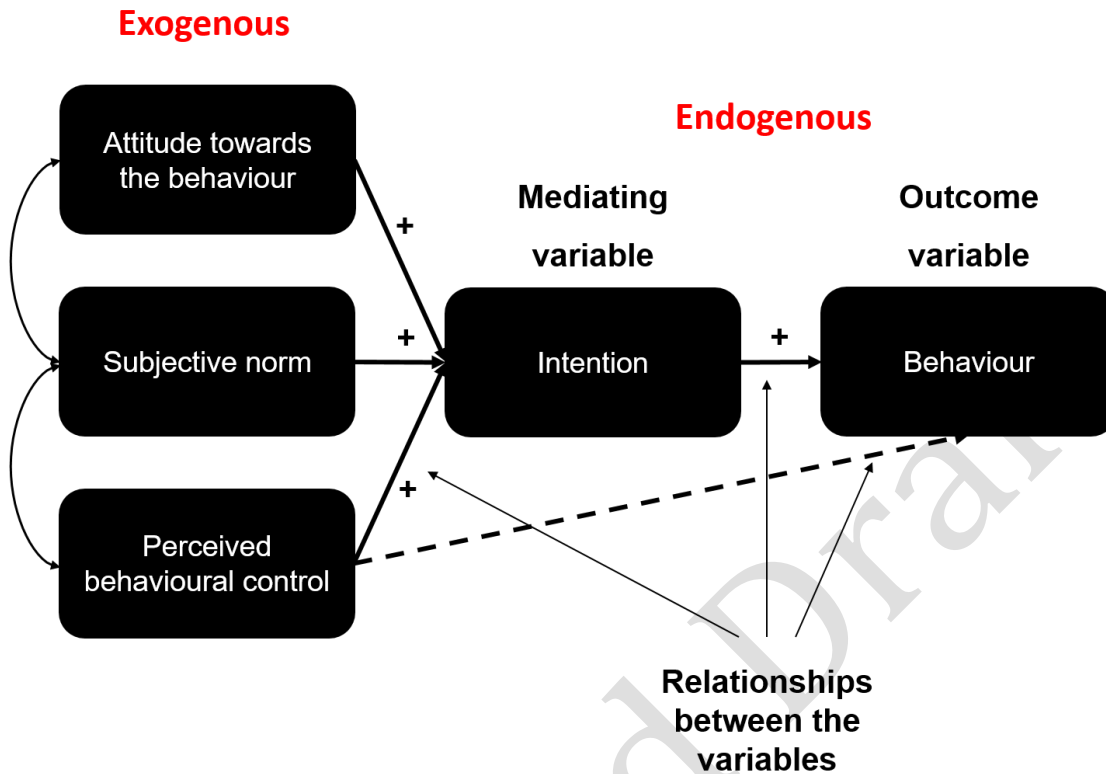


Figure 2: Delineating the model building blocks of the simplified Theory of Planned Behaviour (Ajzen 1991)

**Equations as representations:** A very familiar example of an equation used to represent the relationship between independent and dependent variables is the  $y = a + bx$  equation derived through regression analysis. For multivariate models this formula becomes  $y = a + b_1x_1 + b_2x_2 \dots b_nx_n$ . Where  $y$  is the value of the dependent variable, ' $a$ ' is a constant (i.e., the y-intercept), ' $b_1$ ' is the regression coefficient for variable ' $x_1$ ' (the first independent variable) and ' $b_2$ ' is the regression coefficient for variable ' $x_2$ ' (the second independent variable). We can continue to add ' $b_nx_n$ ' terms for ' $n$ ' variables. For example, we could derive a model to understand influences on consumer earnings and represent this model as an equation,  $y = -5.504 + 0.495x_1 + 0.210x_2$ , where  $y$  represents monthly income,  $x_1$  represents years in education and  $x_2$  represents years in work. We could work out the value of a person's monthly income using this formula (substituting the known values of  $x_1$  and  $x_2$ ). (Note that for completeness we should also include an error term in this formula, not indicated here.)

**Other representations:** Whilst visual (path models) and equations representations are popular formats, you will come across other formats. For example, graphical or chart representations such as the Kano Model (figure 3). This model identifies five categories of customer requirements and illustrates their different relationships with customer satisfaction. The idea being that if you know how your customer requirements affect their

satisfaction then this informs development and management of products. A specially devised assessment tool accompanies this model (Kano Survey), the results of which can be directly plotted onto a graph.

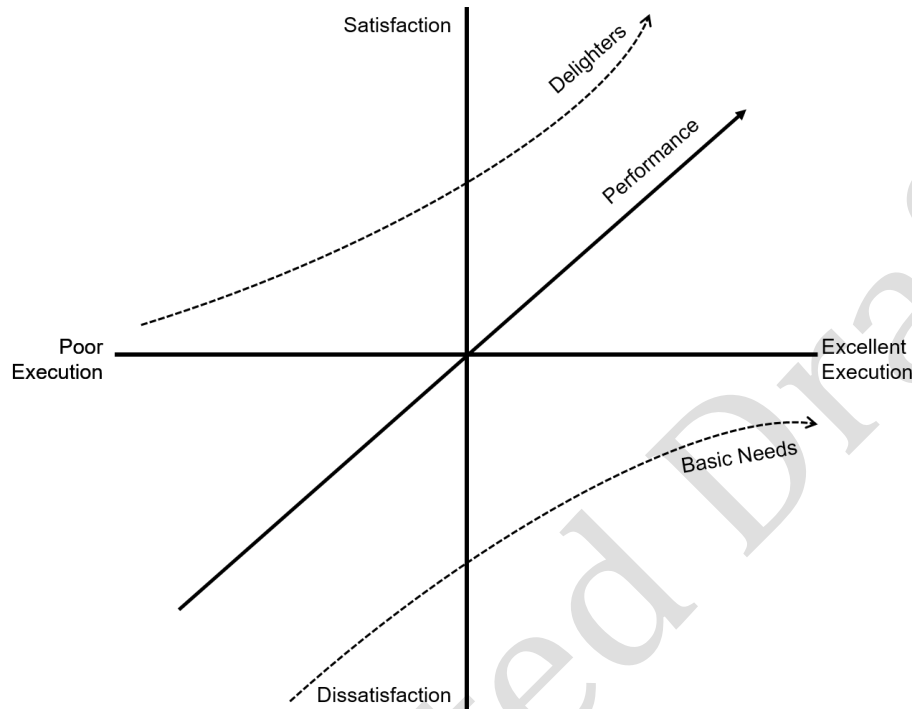


Figure 3: A simplified version of the Kano model (1988)

## 4.2 Constructs and variables

Theories are based on identifying abstract constructs, for example, perceived behavioural control, self-efficacy, customer satisfaction and behavioural intentions. These constructs are often referred to as latent. That is, they cannot be observed or measured directly. In modelling terms, we use the terms **unobserved variable** or **latent constructs** to refer to a construct that is theorised to exist, but is not directly observable. With reference to figure 2, examples of latent constructs are attitude towards the behaviour, subjective norm, perceived behavioural control, and intention. We, therefore, in our analysis approximate the latent constructs by using observable or measurable indicator variables. These are called the **observed variables**, sometimes referred to as **manifest** or **indicator** variables. This type of data is collected from consumers through data collection methods such as questionnaires and observations in experiments. Of course, we can incorporate directly observable variables into our model too, such as salary, age and so forth.

Furthermore, partly when referring to confirmatory analyses, we can distinguish between exogenous and endogenous variables based on their role within the model. **Exogenous** variables refer to the latent equivalent of independent variables that are not influenced by



other variables in the model. They act as predictors of other variables in the model, but are not predicted by other variables. **Endogenous** variables are the latent equivalent to dependent variables, they are affected by other variables in the model. These variables are the 'outcome' in at least one contributory relationship. As shown on figure 2, the endogenous variables are both those based in the 'middle' of the model, e.g., intention is influenced by the three preceding variables, as well as incorporating the outcome variable, e.g., a specific behaviour.

### 4.3 Measuring Model Variables

When representing constructs within our model using observed variables, what we are doing is translating our abstract concepts into measurable (and therefore testable) variables through a process of **operationalization**. Often latent constructs are complex and cannot be measured using a single indicator. Instead, we measure these constructs, indirectly, using multiple indicators (e.g., items on a questionnaire). This requires that we clearly define how we are going to operationalize (put into practice) and hence measure our model variables. Whilst it is beyond the scope of this chapter to discuss the construction of measures (or scales), the importance in terms of contemporary modelling stems from advances in analytical techniques. The development of analytics such as SEM allows us to more readily include directly the measurement items used for each latent construct within our modelling (i.e., not the scale total as with regression techniques), and as such, we can incorporate **measurement error** directly into the estimation process. This leads to better model specification. **Error** refers to the unexplained variance within a model, which we must acknowledge and account for.

We can further distinguish between formative and reflective indicator variables used to measure the unobserved variables (latent construct). With **formative** indicator variables the arrows between the unobserved variable and the observed indicator variables go from the observed indicator variables to the unobserved variable. This indicates that the observed indicator variables 'cause' the unobserved variable. Error in this case can be defined as the inability of the observed indicators to fully explain the unobserved variable. Often a set of formative indicators used to measure an unobserved variable is called an **index**. (See Hair et al., 2016, for an important discussion on the distinction between composite and causal indicators in CB-SEM and PLS-SEM.) With **reflective** indicator variables the arrows between the unobserved variable and the measurement indicators go from the unobserved variable to the observed indicator variables. This indicates that the unobserved variable 'causes' the observed indicator variables. A change in the unobserved variable would bring about a change in all of the observed indicator variables. Error in this circumstance can be defined as the inability of the unobserved variable to fully explain the observed indicators. Often a set of reflective indicators used to measure an unobserved variable is called a **scale**.

### 4.4 Representing Relationships

We have emphasised that the generic goal of modelling is about understanding relationship complexity. Referring to path models, on figure 2 we have represented the theorised relationships between the variables using arrows. We can distinguish between different types of relationships within a model. **Recursive** relationships (indicated on

figure 2 with an arrow going in one direction between two variables) are where one variable X is theorised to impact on another variable Y (therefore the arrow points from X to Y). Conversely, a **non-recursive** relationship, which would be represented by a double-headed arrow, indicates a relationship where X influences Y and Y influences X. A **correlational** relationship (indicated on figure 2 by a curved double-ended arrow) does not indicate any specific direction of influence.

### 4.5 Variable Roles

Variables also have different specific roles within a model. For example, the exogenous variables that we met earlier have a predictive or explanatory role within the model. On the other hand, endogenous variables can have both a predictive/explanatory role and can themselves have a dependent role on antecedent variables.

As such, endogenous variables can take on the role of mediation. A **mediating** variable is a variable that intervenes between two other related variables. That is, the mediating variable explains (facilitates) the relationship between the two other variables. On figure 2, intention is theorised as a mediating variable. For example, the relationship between subjective norm (X) and behaviour (Y) is facilitated by intention (M). **Parallel** mediation is where a number of variables ( $M_1$ ,  $M_2$ , etc.) mediate the relationships between X and Y separately (i.e., they are unrelated) to each other. So the mediating variables ( $M_1$ ,  $M_2$ , etc.) would all be influenced by X and then separately influence Y. **Serial** mediation can be thought of as a 'causal chain', such that, X influences  $M_1$ , which influences  $M_2$ , which influences Y.

Other variables may take peripheral, but nonetheless vital, roles within the model. Most important of these is moderation. A **moderating** variable is a variable that changes the relationship between two other related variables. That is, the existence of a moderator means that the relationship between two variables changes with the level of the moderating variable; i.e., the moderating variable influences either the strength and/or direction of the relationship. We refer to **moderated mediation**, where one variable modifies the effect of a mediating variable.

## 5. Two Examples of Modelling

The two cases provided below are taken from real projects, although as indicated in the footnotes the data has been simplified or modified to allow illustration of the modelling, whilst protecting confidentiality in the cases. The purpose of these cases is to demonstrate the varied contexts and methods that can be classed under the broad modelling label.

### 5.1 Case 1 – Testing the UTAUT2 Model

Case 1 provides an example of dependence techniques. **Dependence** techniques are where we use of a set of (independent) variables to help describe, explain, and/or predict one or more (dependent) variables. Typical analytical techniques include Regression, Discriminant Analysis, Conjoint Analysis, Partial Least Squares-SEM (PLS-SEM), and Structural Equations Modelling (SEM). Further, macros such as PROCESS have been specifically designed to test mediation and moderation.

### ***Problem Identification & Context***

A national health organisation wants to understand why consumers (in this case patients with specific conditions) would use the internet as a means for self-managing their health. The goal is to identify the factors that could be targeted to encourage more use of the internet to empower patients.

### ***Context & Sample***

A market research company collected data from patients (N=623, 53.3% female, average age = 44yrs, s.d. = 15yrs) with specified conditions using a structured questionnaire. The questionnaire was composed of a number of scales designed to measure the constructs within the UTAUT2 (Venkatesh et al. 2012) and use of the internet for self-management of health. The simplified model for testing is presented in figure 4.

### ***Analysis***

Following a Confirmatory Factor Analysis (CFA) to confirm the reliability and validity of the measurements, Structural Equations Modelling (SEM) was used to estimate the model parameters. The AMOS software was used in this case.

### ***What can we say about testing the model?***

Overall, we can observe that as hypothesised all three (exogenous) variables have a significant relationship with intention. Effort Expectancy ( $\beta=0.69$ ,  $p<0.01$ ) has the strongest relationship with intention, followed by Facilitating Conditions ( $\beta=0.51$ ,  $p<0.01$ ) and then Performance Expectancy ( $\beta=0.40$ ,  $p<0.01$ ). All three relationships are positive, indicating that as perceptions of Performance Expectancy, Effort Expectancy and Facilitating Conditions become more positive so intention to use the internet for health purposes increases. Further, in its mediating role, intention has a significant effect on use ( $\beta=0.41$ ,  $p<0.01$ ), such that increasing intention to use is associated with higher actual use of the internet for health purposes.

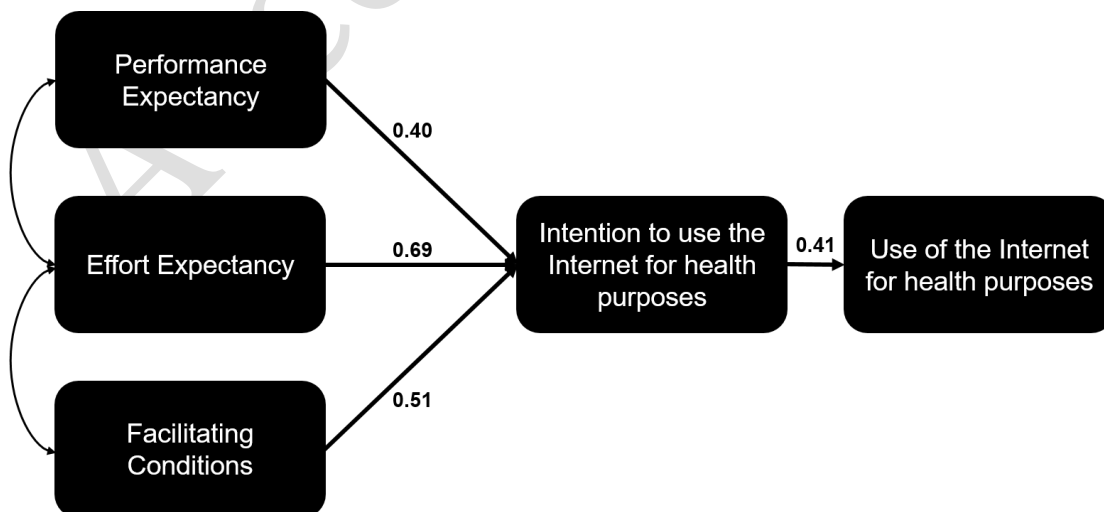


Figure 4: A simplified UTAUT2 model with standardised coefficients

Apart from being able to describe the relationships, we can also assess how good our model is at reproducing our test dataset. SEM does not have one statistical test that we can use to test the strength of the model's predictions. Instead we use multiple goodness of fit indices. Usually, this includes the  $\chi^2/df$  (referred to as CMIN in the AMOS software), a goodness of fit index (e.g., GFI, CFI, NFI, TLI) and a badness of fit index (e.g., RMSEA). We can evaluate our model by looking at these indicators of goodness of fit. We observe that CMIN/df = 3.206 [values between 2-4 are acceptable], CFI = 0.957 [values above 0.9 are acceptable], and RMSEA = 0.067 [values below 0.08 are acceptable]<sup>1</sup>. From this we can conclude that our model is an acceptable fit for our data. (See Hair et al. 2013 for an excellent discussion on the use of these indices in SEM.)

On a note of caution - remember that establishing a good model fit (or establishing that a model is not 'bad') does not mean that this is the only model that could explain the patterns in our data, only that it is a good fit for reproducing our dataset. We can, of course, go on to refine our model by exploring (i) possible moderators (e.g., age and gender) and (ii) potential other variables that may impact on use. On a related note, there is also some controversy about the goodness (or badness) of fit indices and their use. (See, for example, Kenny et al. (2015) for a discussion on the use of RMSEA.) You will find that there are many available indices and that different disciplines favour different indices. You should be comfortable that you understand those relevant to your particular area, but always be aware of challenges to these.

### 5.2 Case 2 - Segmentation and Profiling

Case 2 provides an example of Interdependence technique. **Interdependence** techniques are where we are focus on understanding the underlying structure of data, assessing interdependence without any associated dependence relationships. That is, the goal is to understand how variables can be usefully grouped together. For example, consumer populations are often divided into subgroups sharing characteristics as in target markets. Typical analytical techniques include Cluster Analysis, Factor Analysis and Multidimensional Scaling.

#### **Problem Identification & Context**

A global commercial organisation wants to develop a diagnostic tool for profiling its business partners within a specific geographical region. The goal is to prioritize actionable strategies for those partners who have the highest likelihood of engaging with and climbing up the ladder of the organisation's loyalty programme.

#### **Context & Sample**

The dataset comprises transactional behaviours (655,000) over a one year period, encompassing 98,000 partners. The total number of products sold over the year is 2,300,000, with a value of \$600,000,000<sup>2</sup>. Notice that this data is all observed, using

---

<sup>1</sup> The figures reported here are for illustration purposes only.

<sup>2</sup> The actual sales figures have been altered for confidentiality purposes - but the relative scaling is preserved.

automatic capture of transactions, rather than self-reported as in case 1.

### ***Analysis***

Providing an example of an application of interdependence techniques, cluster analysis is used as a way to identify the underlying patterns within the data that can help reveal key segments. These identified segments are used in this case as a basis for defining the model loyalty programme member. Following this stage, propensity scoring is used to match non-member cases in the dataset to the identified model segment. In this way the company can not only understand the segments within their dataset but also apply actionable strategies for enrolling more members in its loyalty programme.

### ***What can we say about the cluster identification?***

In stage 1, our analysis suggests that 3 segments will be useful as a way of succinctly understanding our sample (table 1). There 3 groups can be identified by a combination of their measurements on 4 key indicator variables (following exploration with a wider dataset).

The company identified segment 3 as their model **elite** profile. However, it was recognised that this elite profile was unrealistic for a large section of their partners to achieve. As such, segment 2 was identified as the model **target** profile. That is, this was the profile that they wanted to facilitate their partners to achieve.

<b>3 Main Segments</b>	<b>Av. No. Trans- actions</b>	<b>Av. No. Units Purchased</b>	<b>Av. Total Sales (U.S. \$)</b>	<b>Av. No. of Portfolio Divisions</b>
<b>1: Low value, low transaction frequency, low variety</b>	3	6	1,500.00	1
<b>2: Moderate value, moderate transaction frequency, moderate variety</b>	8	21	6,000.00	2
<b>3: High value, high transaction frequency, high variety</b>	88	380	120,000.00	4

*Table 1: Resultant 3 segments<sup>3</sup>*

### ***How does this relate to propensity scoring?***

In stage 2, propensity scoring was undertaken to assess which of the non-members in the dataset were most closely matched to the segment 2 model target profile. Table 2 shows the number of matched cases according to the propensity scoring.

<sup>3</sup> The figures reported here are for illustration purposes only – for confidentiality purposes the actual figures have been altered, but the relative scaling is preserved.

	Propensity score range	Number of matched cases	Matched variables
<b>Best matching companies</b>	76-100%	n=20	Good match across all 4 variables
<b>The next most attractive segment</b>	51-75%	n=70	Good match on portfolio divisions
<b>Middle of the road</b>	26-50%	n=350	Good match on average total sales
<b>Lowest engagement</b>	0-25%	n=95,000	Low value transactions across all variables

Table 2: Cases matched to model group 2 by propensity<sup>4</sup>

From this the company were able to develop a prioritised target list of cases as part of their strategy for widening their loyalty programme membership. They started by recruiting the companies in the best matching group indicated above.

## 6. Troubleshooting Tips

When undertaking any kind of modelling you will most likely encounter problems – even in the best design. Six common problems are identified with troubleshooting tips.

### Tip 1: There is always error!

If we want to increase confidence in our work and its applicability then we need to control error as much as we can. Error can occur at any stage of a project, but thoughtful project management can go a long way toward minimising this. When planning your project think about the potential sources of error and how these might be controlled right from the problem formulation stage. Common sources of error are:

- Neglecting to take account of non-response bias, that is, do not forget that it is as important to consider who is not in your study as well as who is. For example, is an important group not represented in your study? Ensure that you include a consideration of the non-response group in your write-up.
- Difficult to control but nevertheless important to consider is the impact of historical events during data collection. Has anything occurred during data collection that could have impacted on or biased the data? In particular, did this affect one group or participants or all?
- Poor data collection practices, such as inadvertently influencing a participant in the study or not mitigating against missing data, which can invalidate your dataset and any subsequent analysis.
- Incorrect application of analyses, especially working outside the limitations of analytical methods (e.g., not observing good data screening practice).

<sup>4</sup> The figures reported here are for illustration purposes only – for confidentiality purposes the actual figures have been altered, but the relative scaling is preserved.

**Tip 2: Causality is difficult to prove!**

**Cross-sectional** design, data collected at one point in time from a sample selected to represent the population of interest, is commonplace in consumer psychology. However, it is important to remember that the aim is to describe a population or document and test differences in a subset of the population at one point in time. As such, you are limited in what you can conclude about causality. **Longitudinal** design, where the aim is to examine a dynamic process that involves change over time and understand the sources and consequences of a phenomenon, is a desirable design but often not possible through resource constraints. The underlying principle is that we measure some dimensions of interest of a given entity before and after an intervening phenomenon to determine whether or not the phenomenon has some effects. As such this gives greater confidence for causal inferences than cross-sectional designs. Remember that demonstrating a statistical fit for your proposed model does not immediately ‘prove’ your causal assumptions about the relationships between variables. Instead it makes them more plausible. **Replication** studies, that is, repeating a study in another sample(s), or ruling out **alternative** models (often within the same study), that is, comparing a different model to your proposed model, provide further evidence and confidence in a model.

**Tip 3: Larger does not always mean better!**

Researchers can be tempted to collect as much data as possible. But size of the dataset is not everything. Quality of the dataset should be the first consideration with a focus on collecting pertinent and well-measured data. With regard to sample size, some analyses can be too sensitive with larger samples. Before you start any data collection, consider the analysis to be used to test the model and what data is needed to test that model using that analysis. Ensure that you design your data collection techniques so that they are adequate for the purposes of that analysis. Also, consider the necessary size of the dataset for the analysis to be used. Some considerations that help you to determine sample size are: the multivariate distribution of the data; the estimation technique to be used (e.g., Maximum Likelihood Estimation); and, the model complexity (and the required number of indicators per construct where relevant). Most reference books on specific analytical techniques will provide guidance on sample sizes and/or variable to N ratios. With regard to the number of variables included in the model. Your aim should be to be reasonably **parsimonious**. That is, that you should identify the most pertinent variables in your model. Simply adding variables may increase your goodness of fit, but not necessarily the theoretical or practical relevance of your model. Do not allow yourself to be driven by achieving better statistical outcomes – instead be driven by achieving a better quality, meaningful model.

**Tip 4: Reliability does not mean validity!**

It is of course vitally important that you develop and/or use measures that are both reliable and valid. If you are using pre-existing scales then check the reported reliability and validity, but do not just rely on these, check for yourself with your own data. Pay special attention when working in cross-cultural contexts (see tip 5). If you are developing your own measures then remember that you will need time to develop these

and to establish their reliability and validity. It is not sufficient to demonstrate reliability of your measures – you can have a perfectly reliable measure that is not valid (i.e., does not measure what you think it measures). As such, you must also establish validity. (*See Oppenheim 1992 for a classic text on scale development.*) Remember that even methods that directly observe behaviours need to be validated (including automated data capture).

### **Tip 5: Context and culture matter!**

Models cannot be separated from the context(s) or culture(s) within which they were developed. When applying a model to another context or culture remember to consider how this may impact on or challenge the underlying assumptions of the model. You need to consider a number of cross-cultural/contextual issues:

- Do the underlying constructs hold for the new context/culture – that is, do they make sense, do they need to be adapted, or do they need to be totally rethought?
- Do the measures used to test the model hold for the new context/culture? Consider not just language translations, but also semantics and use of scales or response formats (and the use of appropriate stimuli in different cultures).
- Do you need to adapt the modelling process? For example, access to participants, data management issues, literacy levels in the focal country/population are all important considerations.

### **Tip 6: Equivalence is illusive!**

One of the things that you may wish to establish is whether your model is equivalent or different across groups (e.g., gender/age groups, teams/departments, cultures/countries). Cross-Validation is an attempt to reproduce the results found in one sample (or population) using data from a different sample – that is, to establish equivalence or invariance across groups. However, it is often the case that you will not achieve full equivalence (*see for example, Malhotra and McCort, 2001, for a comparison of behavioural intentions models across cultures*). You must therefore decide what level of equivalence is adequate to address your objectives. For example, a partial cross-validation using SEM might establish that factor loadings are equivalent; this is typically considered adequate. Tight cross-validation – whilst ideal is rare – where you establish the equivalence of factor loadings, inter-factor covariances, and error variances. (*See Hair et al. 2013 for a fuller discussion on statistical equivalence and Craig and Douglas for a fuller discussion on different methods in cross-cultural research.*)

## **7. Future Directions of Modelling in Consumer Psychology**

We have focused on standard methods used in contemporary modelling in consumer psychology. However, this is a world of constant innovation, not only in terms of developments in methods, analytics and software, but in the problems faced within the world of consumer psychology that need increasingly sophisticated approaches to resolve. There are three identifiable trends in modelling.

### ***A New Era for Bayesian Modelling?***

A recent review (van de Schoot et al. 2017) has shown a clear upward trend in the use of Bayesian approaches within Psychology. Although not a new technique, the accessibility



of this approach has been enhanced by both an increasing supply of communications on methods and developments in software applications. With specific relevance to consumer psychology, there has been increased use of Bayesian approaches within the areas of cognition, personality and neuropsychology, alongside the contexts of health, education and development (Depaoli et al. 2017, van de Schoot et al. 2017). There are parallel developments from a methods perspective. There are equivalent Bayesian approaches to the popular frequentist approaches, e.g., Bayesian SEM, and there is a trend in studies demonstrating Bayesian approaches as credible alternative estimation procedures. Simulation studies focus on the development of and comparison between alternative Bayesian methods (van de Schoot et al. 2017).

Of the reasons for using Bayesian approaches identified by van de Schoot et al. (2017), those applicable to consumer psychologists are (1) the intuitive appeal of explicitly using prior knowledge to estimate models, and (2) the ability of Bayesian models to handle more complex models. Examples of the application of diverse Bayesian approaches to consumer psychology demonstrate the potential of these approaches to address contemporary consumer problems. Lee et al. (2011) use Bayesian approaches to forecast the future digital video-player market taking into account consumer preferences. Jerath et al. (2014) use Bayesian approaches to provide insight into consumers' behaviours in online searching tasks. Jang et al. (2016) predict consumer spending patterns and share of wallet to inform consumer targeting strategies. Swani et al. (2017) demonstrate the propensity of users to popularise (i.e., like and/or comment) brand-related posts on social media. Marinova et al. (2017) investigate how physicians make trade-offs between medical and non-medical information to determine a patient's access to life-enhancing products. Lin et al. (2017) use Bayesian approaches to investigate how to develop an understanding of a patient's risk of future adverse health events to support clinical decision making in personalised and preventive care.

Whilst there may be a divide between those favouring the frequentist approaches or Bayesian approaches, there is potential for applying the methods in a complementary fashion. This will allow advances in consumer psychology that are based on robust development and tests of models and address the complexities faced by consumers in the real world.

### ***A Kickstart for Predictive Modelling?***

Another evident trend in modelling is a resurgence in the use of PLS-SEM (or PLS Path Modelling). Despite evolutions in this technique, there was a paucity of mainstream supporting texts until Hair et al. (2013). Emphasising the rapid developments in PLS-SEM in recent years, the second version closed followed the 2013 edition to document these developments (Hair et al. 2016). Whilst this analysis is ideally placed as a model development tool (Morard and Simonin 2016), it has helped to kickstart studies in the marginalised area of predictive modelling (see section 3.1 above). PLS-SEM supports predictive modelling in complex situations, especially where there is little theory. Beyond this PLS-SEM can cope with modelling when faced with problems such as the issue of small consumer populations or samples (Hair et al. 2014). As such, it is a useful management tool to also provide insight into model optimisation, free from theoretical

and statistical assumptions (Morard and Simonin 2016).

Contemporary examples of consumer psychology relevant studies involving the use of PLS-SEM as a predictive modelling tool can be found across a diverse range of fields. Ayeh et al. (2013) developed a predictive model of intention to use consumer-generated media for travel planning. Lehto and Oinas-Kukkonen (2015) predicted continued engagement in a weightloss programme. Wakefield (2016) used fan passion to predict attendance, media consumption and social media engagement. Schubing et al. (2016) used PLS-SEM for predictive modelling of the diffusion process in consumer networks.

### ***A Big Data Push for Modelling?***

The increasing availability of consumer data has driven sophisticated advances in data mining techniques and software, and enhanced computing power to handle storage, management and analysis of such data. These developments have resulted in more accessibility to 'big data' for modelling in consumer psychology to non-computer science specialists. Cheung and Jah (2016) argue that psychologists may have historically lacked the programming or computational skills for data mining, but that they bring valuable insight into the interpretation of such analyses. Particularly, in terms of understanding and modelling of consumer behaviours and methods of assessing the reliability and validity of data. Indeed, Seitz et al. (2017) argue that the strength of simulation modelling is the ability to achieve parsimony, but this algorithmic approach needs to be embedded within an understanding of the psychology of consumers to avoid reductionism.

The volume, variety, velocity and veracity of available data is driving new areas of research and the discovery of new models and analysis (Cheung and Jah 2016, Goldstone and Lupyan 2016). The variety and volume of data offers the opportunity for more nuanced versions of existing models. In particular, the ability to locate and test models of behaviour in diverse online (new) cultural and social environments (e.g., consumer engagement in online brand communities, Brodie et al. 2013). These environments may also be more suited to modelling of sensitive behaviours, and offer up new ways of establishing the ecological validity of the results from laboratory experiments (Goldstone and Lupyan 2016). The velocity and veracity of available data means that we can not only collect data more rapidly but also have access to more observable naturally occurring data. Meaning that consumer psychologists can now be more agile in their response to developing solutions to managerial and societal problems. Indeed, technological advances mean that some of these phenomena and/or data did not exist before, opening up entirely new areas, and connections between consumers, brands, services and practitioners have increased exponentially. For example, patterns of online collaboration and networks, reactions to regional events on a global scale, and new forms of celebrity are all areas of interest to consumer psychologists (Goldstone and Lupyan 2016).

These modelling efforts are informed and enhanced by advances in the development of statistical algorithms and artificial intelligence (AI). Opening up avenues for delivering services based on automation (e.g., the use of avatars, Kohler et al. 2011), aiding

consumers in making decisions (e.g., recommendation systems, Yin et al. 2017), and informing how companies approach forecasting and prediction of consumer behaviour (e.g., neural networks, Greene et al. 2017). Moving forward, we are likely to see more and more consumer psychologists working within multidisciplinary teams (e.g., alongside informatics, computer scientists and AI experts) in order to push the boundaries of knowledge within our own discipline. As such, this emphasises the wider process of moving from data mining, through to modelling, through to application (often management).

### 8. Software

There are a plethora of paid-for and free-to-download software packages to help you with your analysis. Here is a selection (and certainly not an exhaustive list):

1. General Packages
  - a. AMOS (<https://www.ibm.com/uk-en/marketplace/structural-equation-modeling-sem>)
  - b. SAS ([https://www.sas.com/en\\_in/software/stat.html](https://www.sas.com/en_in/software/stat.html))
  - c. SPSS (<https://www.ibm.com/analytics/data-science/predictive-analytics/spss-statistical-software>)
  - d. Process (<http://www.processmacro.org/index.html>)
  - e. R (<https://www.r-project.org/about.html>)
2. PLS-SEM
  - a. SmartPLS 3 (<http://www.smartpls.com>)
  - b. PLS Graph (<http://www.statisticssolutions.com/pls-graph-software/>)
  - c. Also available with options in SAS, SPSS and R
3. Bayesian Approaches
  - a. Win BUGS (Lunn et al. 2000)
  - b. MPlus (Muethen and Muthen 2015)
  - c. SAS, SPSS and R offer different options
  - d. Other software are constantly being developed

### 9. Summary

Modelling is an exciting area of consumer psychology, with application to many problems and contexts. We have covered the founding principles and objectives of the modelling process, which have remained largely unchanged over the course of time. What has changed are the constant innovations in methodologies (especially analyses) and software development that keep pushing the boundaries of modelling. These developments have given rise to some interesting opportunities to work in multidisciplinary teams (especially around exploiting big data in a meaningful way), and opening up of new and innovative areas of research in understanding the consumer.

## 10. Student exercises

These exercises can be applied to your specific area of interest in consumer psychology. You can treat the exercises as standalone tasks or you can combine them to apply to a set of studies. These exercises will help you in your own modelling studies (perhaps in a dissertation or similar). The first exercise asks you to consider your own phenomenon of interest and think about this in modelling terms. The next three exercises ask you to examine the approaches that other studies have taken in detail and to critique these approaches. The final exercise asks you to focus on cross-cultural issues.

1. Choose a phenomenon related to consumer psychology that is in your area of interest. Think about the variables that could help you to understand the phenomenon. Which of these variables would you identify as the dependent variable? How would you organise these variables into a model? Would you identify mediating variable(s)? Are there any variables that could act as a moderator? Do you know enough to create a sufficient model or do you need to find out more information? Is exploratory or explanatory research justified?
2. Do a review of published studies in modelling in your area of interest. Now create a table that lists the purpose of the study (exploratory, predictive or explanatory) and the types of model used (e.g., path model, equation etc.). What type of analysis is used in these studies, e.g., Bayesian techniques, SEM, PLS-SEM (including software packages)? Do you notice any commonality in these studies? Is there a dominant approach to modelling and analysis in this area? Do you agree with the approaches taken? If so or if not – why is this?
3. Identify 3 or 4 published studies based on modelling in your area of interest. Create a table. In column one, list the types of variables used. In column two, list what data has been collected to measure these variables. In column three, list the rationale given for use of such measures. In columns 4 and 5 list what you think are the advantages and disadvantages of these measures. Do the authors of the studies pick out the same advantages and disadvantages as you?
4. Identify 3 or 4 published studies on modelling in your area of interest (you could use the same studies in as exercise 3). Now think about the following three categories of error: measurement; field-work; data analysis. From your chosen studies identify what sources of error there may within your chosen studies under these three categories. How might you mitigate against these sources of error?
5. In their article “On improving the conceptual foundations of international marketing research”, Douglas and Craig (2006) state that “*All too frequently, cross-country research begins ....without consideration of the underlying conceptual framework and related constructs and their applicability in other research contexts.*” List the issues from their conceptual framework. Think of examples for each of these issues from your own area of interest. What problems might these issues raise for you in modelling?

### 11. Recommended Reading

Craig, C. S., & Douglas, S. P. (2005). *International marketing research*. Chichester: John Wiley & Sons.

Field, A. (2017). *Discovering statistics using IBM SPSS statistics*. Sage.

Field, A., Miles, J. & Field, Z. (2012). *Discovering statistics using R*. Sage.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2013). *Multivariate Data Analysis*: Pearson Education Limited.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publications.

Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach (2<sup>nd</sup> Edition)*. New York: Guilford.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural equation modeling*, 10(3), 333-351.

Lambert, B. (2018). *A Student's Guide to Bayesian Statistics*. Sage.

Oppenheim, A. N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*, (New Edition). Pinter Publishers; London and New York

Sarstedt, M., Ringle, C. M., & Hair, J. F. (2014). PLS-SEM: Looking Back and Moving Forward. *Long Range Planning*, 3(47), 132-137.

van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2), 217.

### 12. References

Ayeh, J. K., Au, N., & Law, R. (2013). Predicting the intention to use consumer-generated media for travel planning. *Tourism Management*, 35, 132-143.

Brodie, R. J., Ilic, A., Juric, B., & Hollebeek, L. (2013). Consumer engagement in a virtual brand community: An exploratory analysis. *Journal of Business Research*, 66(1), 105-114.

Craig, C. S., & Douglas, S. P. (2005). *International marketing research*. Chichester: John Wiley & Sons.

Cheung, M. W. L., & Jak, S. (2016). *Analyzing big data in psychology: a*

split/analyze/meta-analyze approach. *Frontiers in psychology*, 7.

Depaoli, S., Rus, H. M., Clifton, J. P., van de Schoot, R., & Tiemensma, J. (2017). An introduction to Bayesian statistics in health psychology. *Health psychology review*, 11(3), 248-264.

Douglas, S. P., & Craig, C. S. (2006). On improving the conceptual foundations of international marketing research. *Journal of International Marketing*, 14(1), 1-22.

Goldstone, R. L., & Lupyan, G. (2016). Discovering psychological principles by mining naturally occurring data sets. *Topics in cognitive science*, 8(3), 548-568.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2013). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)*. SAGE Publications.

Hair Jr, J. F., Hult, G. T. M., Ringle, C., & Sarstedt, M. (2016). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Sage Publications.

Hair, Joe Jr, F., Sarstedt, M., Hopkins, L., & Kuppelwieser, V. G. (2014). Partial least squares structural equation modeling (PLS-SEM): An emerging tool in business research. *European Business Review*, 26(2), 106-121.

Hayes, A. F. (2013). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach*. New York: Guilford, 1609182308.

Humphreys, A., & Jen-Hui Wang, R. (2017). Automated Text Analysis for Consumer Research. *Journal of Consumer Research*.

Jang, S., Prasad, A., & Ratchford, B. T. (2016). Consumer spending patterns across firms and categories: Application to the size-and share-of-wallet. *International Journal of Research in Marketing*, 33(1), 123-139.

Jerath, K., Ma, L., & Park, Y. H. (2014). Consumer click behavior at a search engine: The role of keyword popularity. *Journal of Marketing Research*, 51(4), 480-486.

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486-507.

Köhler, C. F., Rohm, A. J., de Ruyter, K., & Wetzels, M. (2011). Return on interactivity: The impact of online agents on newcomer adjustment. *Journal of Marketing*, 75(2), 93-108.

Lee, J., Choi, J. Y. and Cho, Y. (2011), A forecast simulation analysis of the next-generation DVD market based on consumer preference data. *International Journal of Consumer Studies*, 35: 448–457. doi:10.1111/j.1470-6431.2010.00958.x

- Lehto, T., & Oinas-Kukkonen, H. (2015). Explaining and predicting perceived effectiveness and use continuance intention of a behaviour change support system for weight loss. *Behaviour & Information Technology*, 34(2), 176-189.
- Lin, Y. K., Chen, H., Brown, R. A., Li, S. H., & Yang, H. J. (2017). Healthcare Predictive Analytics for Risk Profiling in Chronic Care: A Bayesian Multitask Learning Approach. *MIS Quarterly*, 41(2).
- Malhotra, N. K., & McCort, J. D. (2001). A cross-cultural comparison of behavioral intention models-Theoretical consideration and an empirical investigation. *International Marketing Review*, 18(3), 235-269.
- Marinova, D., Kozlenkova, I. V., Cuttler, L., & Silvers, J. B. (2016). To Prescribe or Not to Prescribe? Consumer Access to Life-Enhancing Products. *Journal of Consumer Research*, 43(5), 806-823.
- Oppenheim, A. N. (1992). *Questionnaire Design, Interviewing and Attitude Measurement*, (New Edition). Pinter Publishers; London and New York
- Schubring, S., Lorscheid, I., Meyer, M., & Ringle, C. M. (2016). The PLS agent: Predictive modeling with PLS-SEM and agent-based simulation. *Journal of Business Research*, 69(10), 4604-4612.
- Seitz, M. J., Templeton, A., Drury, J., Köster, G., & Philippides, A. (2017). Parsimony versus reductionism: how can crowd psychology be introduced into computer simulation?. *Review of General Psychology*, 21(1), 95.
- Swani, K., Milne, G. R., Brown, B. P., Assaf, A. G., & Donthu, N. (2017). What messages to post? Evaluating the popularity of social media communications in business versus consumer markets. *Industrial Marketing Management*, 62, 77-87.
- van de Schoot, R., Winter, S. D., Ryan, O., Zondervan-Zwijnenburg, M., & Depaoli, S. (2017). A systematic review of Bayesian articles in psychology: The last 25 years. *Psychological Methods*, 22(2), 217.
- Venkatesh, V., Thong, J. Y., & Xu, X. (2012). Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. *MIS quarterly*, 157-178.
- Wakefield, K. (2016). Using fan passion to predict attendance, media consumption, and social media behaviors. *Journal of Sport Management*, 30(3), 229-247.